Network building with Cytoscape App queries to the BioGateway 3.0 triple store

Stian Holmås
Vladimir Mironov
Martin Kuiper
Norwegian University of Science and Technology (NTNU)
{stian.holmas; vladimir.n.mironov; mtrkuiper} @gmail.com

Abstract. We have developed the BioGateway App, a Cytoscape plugin that supports the building of a wish list of questions that are translated to SPARQL queries, which are launched against the BioGateway server. We demonstrate the functionality of the BioGateway App both in simple and more complex use cases, taken from our website https://www.biogateway.eu/examples/

Keywords: triple store, cytoscape app, network building.

1. Introduction

1.1. The BioGateway triple store

The BioGateway (BGW) triple store [1, 2] was one of the first major RDF resources available with biological information. The use of SPARQL to query triple stores proved to be one of the major limiting factors for the wider acceptance of the Semantic Web technologies by systems biologists. Therefore, we now reach out to the relatively large user community of the biological network analysis platform Cytoscape through the development of the BioGateway App plugin.

1.2. The BioGateway Cytoscape App

The main feature of the BioGateway App is the Query Builder, which supports the design of queries that are built from definitions of proteins or genes and a relationship to either an ontology term or another protein or gene. By adding additional query parts line by line, increasingly complex and restrictive or inclusive queries can be composed. The Run Query command converts these to native SPARQL queries that are launched against the BioGateway 3.0 SPARQL endpoint [3].

1.3. The biological data in the store

The main information that is subject to the query is obtained from IntAct [4], UniProtKB [5] and the Gene Ontology database [6]. To allow a user a special focus on gene regulation we also included several resources with regulatory relations of transcription factors and their target genes.

2

2. The domo: querying the store

2.1. Autocomplete support for query building

The selection of terms of interest such as genes, proteins, ontology terms and relation types is facilitated by an autocomplete function that is driven by a REST API of a NoSQL database loaded with all the entity names and metadata from the BioGateway server, to allow quick response times. To ensure compatibility between the App and the BioGateway data, upon startup the App fetches an XML-based configuration file from our webserver. This file contains the relation types and their URIs, default settings and the default layout style for BioGateway graphs, as well as available metadata types and the query constraints that this metadata enables, allowing some updates to the App without requiring the user to reinstall it. User preferences set in the BioGateway tab of the Cytoscape Control Panel are stored between sessions, such as the default query constraints related to species, data sources, and additional selection criteria for querying the BioGateway content.

2.2. Query constraints

Next to specifying the results through the definition of restrictive queries, the Query Constraints section of the BioGateway tab in the Control Panel allows additional constraints for specific relations, such as setting a minimum confidence score for Protein-Protein Interactions, as provided by IntAct (Orchard et al. 2013). This control panel also allows the selection of extra types of metadata to be loaded together with the results, but as this may significantly increase query time, this metadata can also be added after the network is complete (Reload Metadata), so that it can be used for filtering and display options.

3. Scope of the demo

We will demonstrate the functionality of the BioGateway App both in simple and more complex use cases, taken from https://www.biogateway.eu/examples/. While BGW 3.0.0 contains only human data, BGW 3.0.1 (release date expected in December 2019) will include data for 25 best-studied eukaryotes.

4. References

1. Antezana, E. et al. (2009) BioGateway: a semantic systems biology tool for the life sciences. BMC Bioinformatics, 10, S11.
2. BioGateway home page, https://www.biogateway.eu/#database
3. BioGateway SPARQL endpoint: www.biogateway.eu/sparql-endpoint/
4. Orchard, S. et al (2013) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res., 42, D358—D363.
5. Gene Ontology Consortium (2018) The Gene Ontology resource: 20 years and still GOing strong. Nucleic Acids Res., 47, D330—D338.
6. Uniprot ref Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).